Journal of Bioresources 10 (1): 07–13, January – June 2023

JOURNAL OF BIORESOURCES

journal webpage: https://jbr.rgu.ac.in

ISSN: 2394-4315 (Print)
ISSN: 2582-2276 (Online)

RESEARCH ARTICLE

DNA barcoding and molecular phylogenetics of Anguilliform fishes of Assam, India

Shamim Rahman[1,*], Amit Kumar[2], Amalesh Dutta[3], Mohan C. Kalita[4]

[1]Department of Zoology, D. C. B. Girls' College, Jorhat-785001, Assam, India.
[2]Centre for Climate Change Studies, Sathyabama Institute of Science and Technology, Chennai-600119, Tamil Nadu, India.
[3]Department of Zoology, Gauhati University, Guwahati-781014, Assam, India.
[4]Department of Biotechnology, Gauhati University, Guwahati-781014, Assam, India.

*Corresponding author: shamim.piku85@gmail.com
Article No.: SRJBR06; Received: 20.04.2023; Reviewed: 10.05.2023; Revised: 20.05.2023; Accepted: 12.06.2023; Published: 30.06.2023
Doi: https://doi.org/10.5281/zenodo.10513601

Abstract

Fishes have rarely been seen through the angle of conservation needs as compared to large strata mammals. Assam is a part of one of the biodiversity hotspots and as it happens to any other species, fishes are also declining from this region, even before their scientific exploration. DNA barcoding has been proven to be a comparatively quicker, but authentic tool for species identification. In current study, DNA barcodes for Anguilliform fishes of Assam have been developed utilizing the partial mitochondrial Cytochrome Oxidase I (COI) gene. The 36 barcodes generated in present study has demonstrated the delineation of 15 morphologically identified species representing monophyletic clusters. Mean intra-specific divergence levels for most of the species were found to be less than 1%, except *Mastacembelus armatus* which exceeded 2%, which has traditionally been considered a threshold for species determination. Excluding *M. armatus,* the maximum conspecific distance and minimum congeneric distance were found at 1.97% and 8.78%, respectively, thus establishing 4.46 fold barcode gap for species-level discrimination. Comparative analysis with secondary data indicates data voids of *Pillaia indica* and probable misidentification of *Anguilla* spp. and *Botia* spp. in public databases; thus it demands deposition of more DNA barcodes and review of morphotaxonomy. Besides the inherent benefit of barcoding in tagging species, it also provides baseline information on the molecular characterization to decipher indications on the phylogenetic significance of lesser studied fishes of Assam. The novelty of this study lies in the *de novo* development of certain DNA barcodes from this region making this study significant for conservationists and field biologists as a whole.

Keywords: Eel, Ichthyos, mtCOI, Taxonomy, Conservation

1. Introduction

The state of Assam is a constituent unit of the Eastern Himalayan Biodiversity region sustaining a wide range of flora and fauna (Groombridge and Jenkin, 1998). Assam has the highest number of wetlands in the North East India along two major river basins: the Brahmaputra and the Barak. A total of 78,438 square km area in these two river basins, including their tributaries (4820 km length in combination), floodplain wetlands and swamps (1.12 lakh ha), ponds and tanks (0.23 lakh ha) and shallow paddy fields (33.45 lakh ha) sustain Indo-Gangetic and small extent of Burmese and Chinese fish fauna (Ponniah and Sarkar, 2000). Previous studies have reported ~300 species of fishes in the water bodies of Assam (Sen, 2000; Bhattacharya et al., 2000; Goswami et al. 2012).

Fish is an integral part of the socio-economic fabric of Assam. It is a cheap and prime source of animal proteins in the state (Saikia and Ahmed, 2012). Eels and eel-shaped fishes, due to their charismatic shape and size and high market values, are of special interest (Leander et al., 2012). Eels and eel-shaped fishes, such as *Anguilla bengalensis*, *Macrognathus aral*, *Macrognathus pancalus* and *Mastacembelus armatus* bear high market prices due to direct consumption, traditional therapeutics as well as ornamental values (Arunachalam and Sankarnarayan, 2000). The word 'eel-shaped fishes' is an umbrella terminology for diverse fish species, and has a history of taxonomic budge within the group due to similar external morphology (Shagufta, 2012). Conventionally, it includes both Anguilliformes and Synbranchiformes, due to similar morphological patterns called 'anguilliformity' (Hossain et al., 2007). Anguilliformity is also exhibited by loaches (Cobitdae) and certain catfishes (Jansen et al., 2006; Shagufta, 2012; Sarkar et al., 2013).

In the last few decades, freshwater fishes in this eco-region have been facing major threats such as habitat loss, habitat fragmentation, anthropogenic activities, untenable fishing, urbanization, and climate change (Deka et al., 2005). As a result, several fish species are reported to be critically endangered, vulnerable, or threatened (Bhattacharya et al., 2000; Gowswami et al., 2012). Thus, assessing freshwater fish diversity is the need of the hour so that an appropriate conservation approach can be undertaken. Identification of fish species is sometimes challenging owing to the high morphological similarities, thus the use of misnomers, erroneous description of the species and misidentification hamper the conservation strategies (Ardura et al., 2010). In addition, there are chances that many species will go extinct before we can establish their existence, due to the rapid rate of species decline (Dubios, 2003).

DNA barcoding and molecular phylogeny is the best choice for the quick and accurate identification of flora and fauna (Hubert et al., 2008). DNA barcode involves the identification of species through the molecular marker, i.e., ~650 bp of mitochondrial cytochrome oxidase I (COI) gene, developed by Hebert et al (2003). This molecular technique offers several advantages, such as the capability to differentiate closely related cryptic species, resolves taxonomic identities, can be performed on a small amount of body tissues and is applicable at every life stage (Ghosh, 2012). Several studies in recent years have utilized DNA barcodes for freshwater fish diversity in India (Lakra et al., 2016; Kundu et al., 2019; Pandey et al., 2020).

Most of the eel-shaped fishes are threatened (IUCN categories) and declining due to over-exploitation and habitat degradation (Anonymous, 2015). Hence, it is imperative to know about the species for framing out conservation and management strategies, which is not possible without having accurate taxonomic information, in hand. Thus, in the present study, we have utilized DNA barcode and molecular phylogeny for accurate identification of eel-shaped fishes, including Anguilliformes and Synbranchiformes, and anguiliformity showing air-breathing catfishes and loaches (Figure 1, Table 1).

Table 1. Eels, eel shaped fish and fishes exhibiting Anguilliformity of Assam in the present study

S.N.	Order	Family	Genus	Species
Anguilliformes and Synbranchiformes				
01	Anguilliformes	Anguillidae	*Anguilla*	*A. bengalensis*
02	Anguilliformes	Ophichthidae	*Pisodonophis*	*P. boro*
03	Synbranchiformes	Mastacembelidae	*Macrognathus*	*M. aral*
04	Synbranchiformes	Mastacembelidae	*Macrognathus*	*M. pancalus*
05	Synbranchiformes	Mastacembelidae	*Mastacembelus*	*M. armatus*
06	Synbranchiformes	Synbranchidae	*Monopterus*	*M. cuchia*
07	Synbranchiformes	Chaudhuridae	*Pillaia*	*P. indica*
Fish exhibiting Anguilliformity				
01	Cypriniformes	Cobitidae	*Botia*	*B. dario*
02	Cypriniformes	Cobitidae	*Botia*	*B. lohachata*
03	Cypriniformes	Cobitidae	*Pangio*	*P. pangia*
04	Cypriniformes	Cobitidae	*Lepidocephalichthys*	*L. guntea*
05	Cypriniformes	Cobitidae	*Lepidocephalichthys*	*L. goalparensis*
06	Cypriniformes	Cobitidae	*Canthophrys*	*C. gongota*
07	Siluriformes	Clariidae	*Clarias*	*C. magur*
08	Siluriformes	Heteropneustidae	*Heteropneustes*	*H. fossilis*

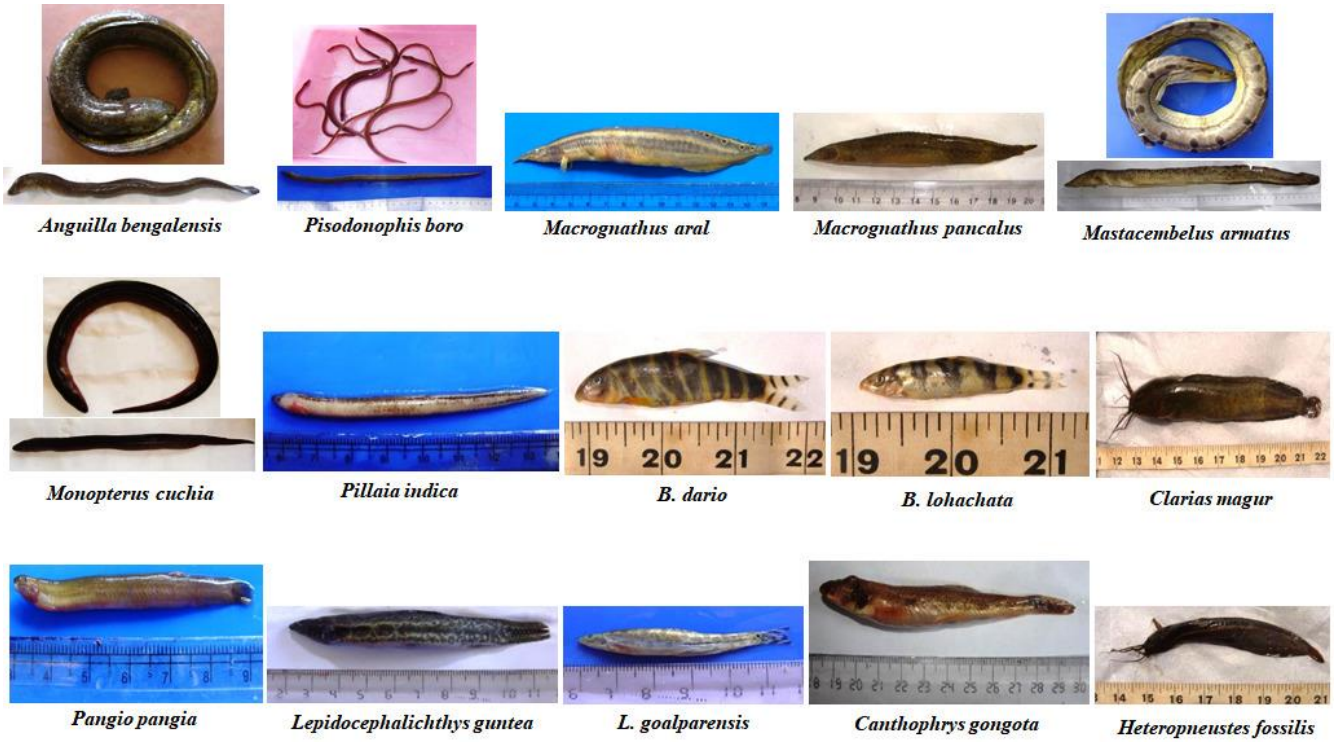

Figure 1. Eels, eel shaped fish and fishes exhibiting Anguilliformity of Assam in the present study

2. Materials and methods

2.1. Sample collection and morphological identification

A total of 36 specimens were collected, either from natural habitats with assistance from local fishermen or the fish market at different locations in Assam (Figure 2, Table 2).
Collected specimens were preserved in -20 °C deep freeze or 95% ethanol after proper cataloguing for molecular studies. Before preservation, samples were photographed for meristic counts and morphological features for identification based on existing taxonomic keys (Talwar and Jhingran, 1991; Viswanath et al., 2007; Jayaram, 2010). Current valid names were cross-verified following Fricke et al (2020).

2.2. DNA extraction, PCR and sequencing

Total genomic DNA (gDNA) was isolated from the muscle tissue of the fish according to Miller et al (1988). The quantity and quality of the gDNA were checked by 1% agarose gel and spectrophotometer (Thermo-Fischer Scientific, USA) for 260/280 values. Approximately 650 bp fragment of mitochondrial gene cytochrome oxidase submit I (COI) was amplified using fish primers: Fish_F: 5′-

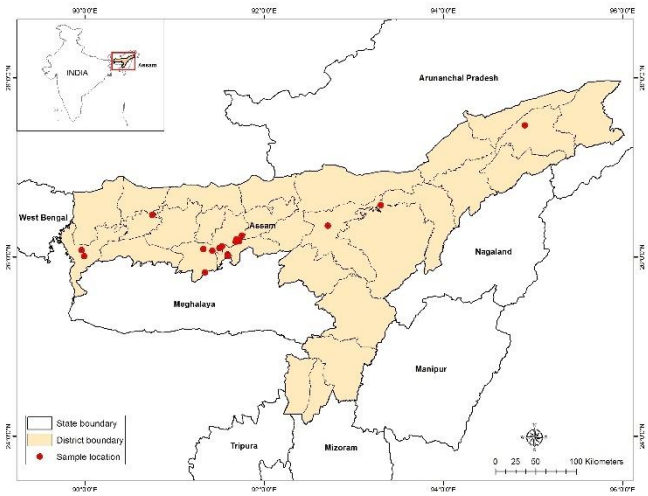

Figure 2. Map of Assam showing sampling locations

Table 2. Sampling stations with coordinates, sampling mode and ecology type.

SN	Sampling station	Coordinates	Mode of sampling	Type of ecology
1	Gauripur	26°05′19″ North 89°58′03″ East	Direct Sampling Market Survey	Lotic -
2	Dhubri	26°01′11″ North 89°59′44″ East	Market Survey	-
3	Nowapara (Near Manas)	26°28′25″ North 90°45′32″ East	Direct Sampling	Lentic
4	Dolani (Near Manas)	26°28′16″ North 90°45′10″ East	Direct Sampling	Lentic
5	Ukium	25°50′22″ North 91°20′33″ East	Direct Sampling	Lentic
6	Gumi	26°05′56″ North 91°19′32″ East	Direct Sampling	Lentic
7	Kukurmara	26°04′20″ North 91°25′44″ East	Market Survey	-
8	Bijoynagar	26°06′05″ North 91°30′21″ East	Market Survey	-
9	Palashbari	26°07′17″ North 91°31′58″ East	Market Survey	-
10	Rani (spot-A)	26°01′48″ North 91°35′41″ East	Direct Sampling	Lentic
11	Rani (spot-B)	26°01′23″ North 91°35′37″ East	Direct Sampling	Lentic
12	Rani (spot-C)	26°01′02″ Noth 91°35′52″ East	Direct Sampling	Lentic
13	North Guwahati (Majgaon)	26°11′16″ North 91°43′24″ East	Direct Sampling	Lentic
14	North Guwahati (Ghorajan)	26°12′17″ North 91°41′46″ East	Direct Sampling	Lotic
15	North Guwahati (Silgurijan)	26°14′51″ North 91°45′22″ East	Direct Sampling	Lotic
16	Pandu	26°10′18″ North 91°41′12″ East	Market Survey	-
17	Nagaon (Kolong)	26°20′59″ North 92°42′49″ East	Market Survey	-
18	Kaziranga	26°35′05″ North 93°18′17″ East	Direct Sampling	Lotic
19	Dibrugarh	27°28′23″ North 94°54′47″ East	Market Survey	-

TCAACCAACCACAAAGACATTGGCAC-3′; Fish_R: 5′-TAGACTTCTGGGTGGCCAAAGAATCA-3′ (Ward et al., 2005). Polymerase Chain Reaction (PCR) was performed in a total of 25 µl containing 12.5 µl 2X master mix, 2 µl each of forward and reverse primer, 1 µl gDNA as a template and remaining nuclease-free water. PCR conditions were as follows: 95 °C for 2 min followed by 35 cycles of denaturation at temperature 95 °C for 30 S, annealing at temperature 55 °C for 30 min, extension at temperature 68 °C for 1 min and final extension at 68 °C for 10 min. The PCR products were purified using HiPura purification kit (HiMidia), before proceeding to bidirectionally sequenced in ABI Genetic Analyzer 3730 (Applied Biosystems Inc., USA) using BigDye® terminator chemistry.

2.3. Sequence analysis

Raw sequence chromatograms were visualized, edited and contigs were prepared using consensus sequences from both the strands in BioEdit v 7.2.5 (Hall, 1999). Sequences obtained in the present study were deposited in NCBI GenBank (KJ946382-KJ946384, KP982886, KX355465- KX355480, KX656908-KX656917, KY172977-KY172981) and BOLD database (AAJ2664 (2), ACP1605 (3), ACA0144 (4), AAF5455 (3), ACS3859 (1), ACA0142 (3), AAF8878 (3), ADE5075 (3), ABU9504 (3), ADE5958 (1), ADC2992 (2), ACC0078 (2), ACY9586 (1), ACC0613 (1), AAM1926 (1), ACR4875 (3). COI gene sequences from the present study were then compared with sequences of related taxa in the public database NCBI GenBank using the BLASTn tool. COI sequences of related taxa were downloaded and aligned with the sequences obtained in the present study in the web version of clustal Omega. The aligned files were then visualized in BioEdit v 7.2.5 (Hall, 1999) for necessary modifications. The aligned sequences were used for maximum likelihood based phylogenetic tree on W-IQ-TREE webserver (Trifinopoulos et al., 2016). The Generalized Time Reversal (GTR) evolutionary model was found to be the best-fit substitution model on this dataset. The calculated parameters were as follows: state frequencies (empirical counts from alignment) pi(A) = 0.2587, pi(C) = 0.2726, pi(G) = 0.1724, pi(T) = 0.2963, substitution rate parameter, A-C: 2.422, A-G: 11.089, A-T: 4.308, C-G: 739, C-T: 17.079, G-T: 1.000. The model of rate heterogeneity was Invariable (0.5011) and Gamma shape alpha (0.996). The robustness of the ML tree was analysed by reiterating the observed data using an ultrafast bootstrap approximation for 1000 generations (Hoang et al., 2017). Number of conserved, variable sites, parsimony informative sites and genetic distance using Kimura-2-parameter (Kimura, 1980) were calculated in MEGA 7 (Kumar et al., 2016).

The DNA Barcode Gap analyses, which examine the distance to the nearest neighbour for each of the species, and sequence composition, were calculated using the BOLD Sequence Analysis Tool.

3. Results

3.1. DNA barcode and species identification

In the present study, DNA barcodes for 36 specimens belonging to 15 morphologically identified species (*Anguilla bengalensis*, n=2; *Pisodonophis boro*, n=3; *Macrognathus aral*, n=4; *Macrognathus pancalus*, n=3; *Mastacembelus armatus*, n=4; *Monopterus cuchia*, n=3; *Pangio pangia*, n=2; *Pillaia indica*, n=3; *Lepicephalichthys guntea*, n=2; *Lepicephalichthys goalparensis,* n=1; *Canthophrys gongota*, n=1; *Botia dario*, n=3; *Botia lohachata*, n=1; *Clarias magur*, n=1; *Heteropneustes fossilis*, n=3) from Assam were generated. Based on NCBI Genbank and BOLD database, a sequence similarity search revealed all the studied morphospecies were closely related to publicly available published database sequences of the

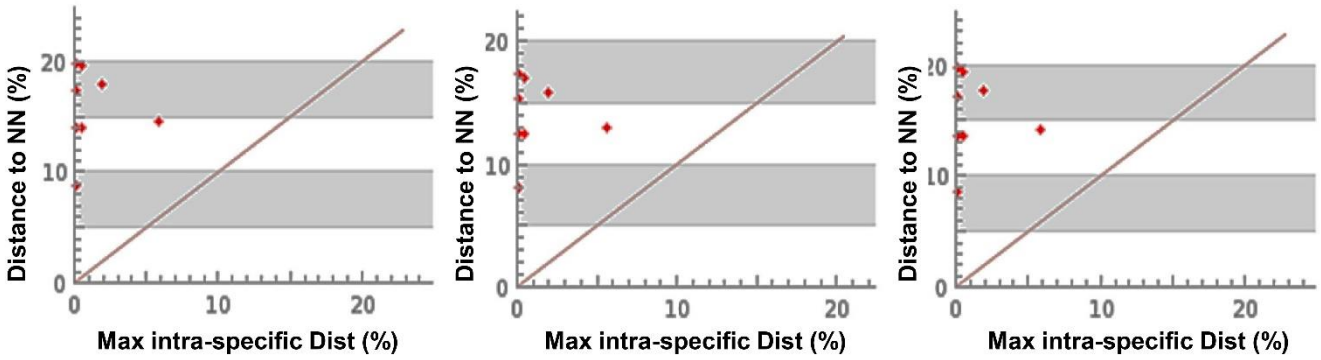

Figure 3. Barcode Gap analysis using three distance methods: a) Kimura-2-parameter; b) Pairwise distance; c) Jukes-Cantor.

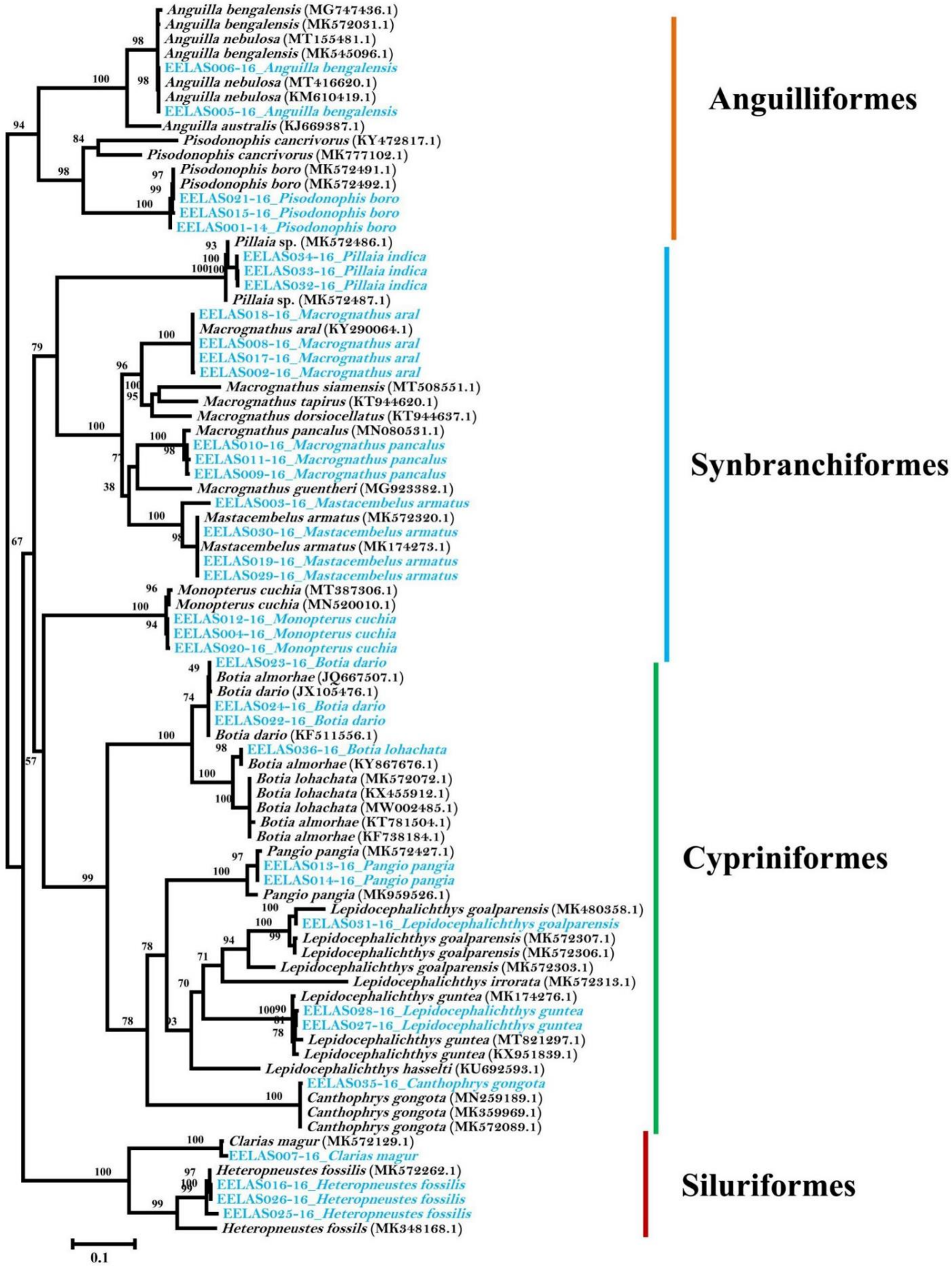

Figure 4. ML-based phylogenetic tree

same species. However, in the case of *A. bengalensis, Botia dario* and

B. lohachata, more than one species of the respective genus was showing high similarity (>98%). *Anguilla bengalensis* specimens were showing high similarity (100%) to A. *bengalensis, A. aguilla, and A. nebullosa*. *Botia lohachata* showed high similarity (>98%) with *B. lohachata, B. rostrata, and B. almorhae*. According to BOLD sequence similarity search, low similarity (82-85%) was obtained for *P. indica* (Table 3).

According to DNA barcode gap analysis using Kimura-2-parameter, Jukes Cantor and P-distance, all the points in the gap analysis were above the slope in the graph between distance to nearest neighbor and intra-specific distance, indicating the existence of a barcode gap (Figure 3). The genetic distances increased with increase of taxon level hierarchically, with a mean, within species distance of 0.83%, within genus distance of 13.36% and within family distance of 16.36% (Table 4). *Mastacembelus armatus* exceeded the conventional intra-specific threshold level in DNA barcode (>2%), while, *P. indica* could not be used due to the absence of reports in the BOLD database. Excluding these species, maximum conspecific distance and minimum congeneric distance were found to be 1.96% and 8.78%, respectively, indicating 4.46 fold barcode gap values.

The nucleotide base composition revealed that out of 640 sites, 375 are conserved, 265 are variable with 11 singletons and 254 parsimony informative sites. It was found that the mean composition of T was highest 29.45%, followed by A (26.16%), C (27.2%), and G (17.19%). Mean GC content was 44.39%, and GC at codon 1, 2, 3 was 55.96%, 43.47%, 34% respectively.

3.2 Phylogenetic analysis

The ML-based phylogenetic tree for the COI gene was constructed with a total of 85 sequences, including 36 sequences generated in the present study. Out of 651 nucleotide sites, 268 sites were parsimony informative sites. The tree topology resulted in 4 major clades belonging to 4 different orders with several subclades with distinct inter-specific separation, supported by high bootstrap values (Figure 4). Clade I consisted of the members of the order Anguilliformes, further separated into two subclades, one each for *A. bengalensis* and *P. boro*. However, no clear distinction was observed between *Anguilla* spp. including *A. bengallensis*, and *A. nebulosa*. Clade II consisted of the order Synbrachiformes, which formed separate subclades for *P. indica, M. aral, M. pancalus, M. armatus* and *M. cuchia*. Clade III consisted of the order Cypriniformes and was further divided into *B. dario, B. lohachata, P. pangia, L. goalparensis, L. guntea*, and *C. gongota*. However, *B. dario* and *B. almorhae* were clustered together. Similarly, *B. lohachata* was found to be associated with *B. almorhae*. Clade IV consisted of the order Siluriformes, split into two subclades, one for *C. magur*, and another for *H. fossilis*.

Table 3. Summary of species identification based on GenBank and BOLD similarity search

Sequences	Species	BLAST- top hit	% similarity	BOLD- top hit	% similarity
EELAS001-14	*Pisodonophis boro*	*P. boro*	99.52%	*P. boro*	>99%
EELAS002-16	*Macrognathus aral*	*M.aral*	99.84%	*M.aral*	>99%
EELAS003-16	*Mastacembelus armatus*	*M.armatus*	100%	*M.armatus*	>99%
EELAS004-16	*Monopterus cuchia*	*M.cuchia*	99.80%	*M.cuchia*	>99%
EELAS005-16	*Anguilla bengalensis*	*A.bengalensis*	100%	*A.bengalensis*	100%
		A.nebulosa	100%	*A.nebulosa*	100%
EELAS006-16	*Anguilla bengalensis*	*A.bengalensis*	100%	*A.bengalensis*	100%
		A.anguilla	100%	*A.anguilla*	100%
		A.nebulosa	100%	*A.nebulosa*	100%
EELAS007-16	*Clarias magur*	*C.magur*	99.84%	*C.magur*	>99%
EELAS008-16	*Macrognathus aral*	*M.aral*	99.54%	*M.aral*	>99%
EELAS009-16	*Macrognathus pancalus*	*M.pancalus*	99.54%	*M.pancalus*	>99%
EELAS010-16	*Macrognathus pancalus*	*M.pancalus*	99.52%	*M.pancalus*	>99%
EELAS011-16	*Macrognathus pancalus*	*M. pancalus*	99.24%	*M. pancalus*	>99%
EELAS012-16	*Monopterus cuchia*	*M.cuchia*	99.30%	*M.cuchia*	>99%
EELAS013-16	*Pangio pangia*	*P.pangia*	100%	*P.pangia*	>99%
EELAS014-16	*Pangio pangia*	*P.pangia*	100%	*P.pangia*	>99%
EELAS015-16	*Pisodonophis boro*	*P. boro*	99.69%	*P. boro*	>99%
EELAS016-16	*Heteropneustes fossilis*	*H. fossilis*	100%	*H. fossilis*	100%
EELAS017-16	*Macrognathus aral*	*M. aral*	99.53%	*M. aral*	>99%
EELAS018-16	*Macrognathus aral*	*M. aral*	99.54%	*M. aral*	>99%
EELAS019-16	*Mastacembelus armatus*	*M. armatus*	100%	*M. armatus*	>99%
EELAS020-16	*Monopterus cuchia*	*M. cuchia*	99.20%	*M. cuchia*	>99%
EELAS021-16	*Pisodonophis boro*	*P. boro*	100%	*P. boro*	>99%
EELAS022-16	*Botia dario*	*B.dario*	100%	*B.dario*	100%
				B. almorhae	100%
EELAS023-16	*Botia dario*	*B.dario*	100%	*B. dario*	100%
				B. almorhae	100%
EELAS024-16	*Botia dario*	*B.dario*	100%	*B. dario*	
				B. almorhae	100%
EELAS025-16	*Heteropneustes fossilis*	*H. fossilis*	99.38%	*H. fossilis*	>98%
EELAS026-16	*Heteropneustes fossilis*	*H. fossilis*	100%	*H. fossilis*	>98%
EELAS027-16	*Lepicephalichthys guntea*	*L. guntea*	100%	*L. guntea*	>99%
EELAS028-16	*Lepicephalichthys guntea*	*L. guntea*	99.80%	*L. guntea*	>99%
EELAS029-16	*Mastacembelus armatus*	*M. armatus*	100%	*M. armatus*	>99%
EELAS030-16	*Mastacembelus armatus*	*M. armatus*	100%	*M. armatus*	>99%
EELAS031-16	*Lepicephalichthys goalparensis*	*L. goalparensis*	99.60%	*L. goalparensis*	>98%
EELAS032-16	*Pillaia indica*	*Pillaia* sp.	98%	*Emmelichthys struhsakeri*	82-85%
EELAS033-16	*Pillaia indica*	*Pillaia* sp.	98.40%	*Emmelichthys struhsakeri*	82-85%
EELAS034-16	*Pillaia indica*	*Pillaia* sp.	98.60%	*Emmelichthys struhsakeri*	82-85%
EELAS035-16	*Canthophrys gongota*	*C.gongota*	99.80%	*C.gongota*	>99%
EELAS036-16	*Botia lohachata*	*B. lohachata*	99.68%	*B. lohachata*	100%
		B. rostrata	98.50%	*B. rostrata*	100%
		B. almorhae	100%	*B. almorhae*	100%
				B.dario	100%

Table 4. Mean genetic distance based on Kimura-2-parameter

Taxonomic level	Min Dist (%)	Mean Dist (%)	Max Dist (%)	SE Dist (%)
Within Species	0	0.83	5.88	0.06
Within Genus	8.78	13.36	14.93	0.15
Within Family	14.54	16.36	22.43	0.06

4. Discussion

In the present study, we have performed barcoding of 36 specimens belonging to 15 species classified among 12 genus, 7 families, and 4 orders. Mean intra-specific divergence level for most of the species were found to be less than 1%, except *M. armatus*. This demonstrates the good species identification capacity of the mitochondrial COI gene for identifying the majority of freshwater anguilliform fishes (Hanzen et al., 2020; Peninal et al., 2017). A 4.6 folds barcode gap, obtained in the present study was found to be similar to that of the previous report (Bhattacharjee et al., 2012). Frequency-based graphical analysis also indicated most of the species, except *M. armatus*, showed low intra-specific divergence and high divergence towards nearest neighbor (within genus or family).

The present study contributed for the first time COI gene sequence of *P. indica* in the BOLD database, as previous matches were insignificant and non-specific with *Emmelichthys struhsakeri*. It indicates the least scientific exploration of this species in terms of molecular biology point of view. In the phylogenetic tree, *P. indica* formed a clade with *Pillaia* sp., recently reported from Bangladesh (Rahman et al., 2019). Based on COI supporting phylogenetic analysis, it was not possible to differentiate between *Anguilla bengalensis* and *A. nebulosa* as they clustered together in one subclade (Figure 4). Also, within the same BIN of BOLD, 17 made *A. bengalensis* and 5 made *A. nebulosa*. Both specific and non-specific match was found in the similarity search. The lesser congeneric genetic divergence was noticed between the two species indicating the possible use of synonym *A. nebulosa* for *A. bengalensis* as mentioned in the past report (Aoyama, 2009). The higher percentage match with *A. bengalensis* (77.27%) in comparison to *A. nebulosa* (22.72%) also corroborates the statement. Talwar and Jhingran (1999) and Jacoby et al (2014) considered *A. nebulosa* as a synonym of *A. bengalensis*. However, few reports, including Eshmeyer's catalogue of fish consider both as separate valid species (Fricke et al., 2020). It requires a thorough study with larger sample size from wider geographical locations to resolve the current confusion between *A. nebulosa* and *A. bengalensis*.

The specimens of *B. dario* showed 100% similarity with both *B. dario* and *B. almorhae* in the BOLD database and also formed the same clade in the phylogenetic tree. Similarly, specimen of *B. lohachata* showed 100% similarity as *B. lohachata, B. almorhae and B. dario* in the BOLD analysis and formed the same clade with *B. almorhae* in the phylogenetic tree (Figure 4). The presence of *B. almorhae* in both subclades indicates a case of misidentification in the public database as the species of *Botia* genus is reported to form a monophyletic group (Dey et al., 2016). A low genetic distance (0-0.4%) among available sequences of four *Botia* spp. including *B. almorhae* (KY867676), *B. histrionic* (KY847870), *B. lohachata* (KY867674), and *B. rostrata* (KY847869), further, indicates a case of misidentification (Kundu et al., 2019). Misidentification of the *Botia* spp. can be possible due to the morphological similarities, changing body patterns through different life stages and sex (Menon, 1992). Though, a thorough study including morphological and molecular analysis is required, in the future, to ascertain this claim.
We also observed a high mean intra-specific genetic divergence for *M. armatus* (2.92%). Morphometric and meristic characters of all specimens were examined to ascertain the species' identity. The high genetic divergence indicates hidden species diversity in the studied region, which could be linked with gene pool fragmentation, as also reported earlier for this species in the Narmada River (Khedkar et al., 2014). Gene flow disruption due to physical and ecological barriers can induce genetic differentiation in fishes (Xing et al., 2020).

DNA barcoding has become a popular tool for identifying fishes and products (Lakra et al., 2011; Pandey et al., 2020). The use of DNA barcoding of identifying eels and eels shaped fishes can offer useful tools for monitoring illegal fish trade and conservation of the species (Hanzen et al., 2020). Example, identification of illegal trade of *A. Anguilla* glass eels from Europe into Hong Kong (Stein et al., 2016); smoked eel species in New Zealand (Smith et al., 2008). Hence, DNA barcoding of the Anguilliform fishes of Assam, which is a part of a biodiversity hotspot, could be helpful in gathering information on stock, harvest and trade at the species level.

5. Conclusion

By developing 36 DNA barcodes and re-performing the phylogenetic organization in light of this new data, we report the DNA barcoding as a useful tool for the identification of eel and eel-shaped fishes in the river systems of Assam. This approach has the potential to strengthen the database for accurate species identification, thereby, helping in formulating sustainable fisheries management practices. Moreover, BOLD database was further enriched by a new dataset on *Pillaia indica*. Also, we highlighted the erroneous labelling of *Botia* spp. and *Anguilla* spp. in the public database, which can hamper the utility of DNA barcoding in species identification. Our results will further pave the way for ongoing and future biodiversity studies of freshwater fishes in the Rivers of Assam.

Acknowledgments

The authors are thankful to the Department of Biotechnology, Government of India sponsored 'Institutional Biotech Hub' of Gauhati University and North Gauhati College for providing the necessary facilities and support to carry out the experiments.

Author's contributions

SR, AD and MCK have conceptualized the problem and designed the experiment. SR has conducted the experiments, analysed the findings and compiled the manuscript. AK analysed the findings and modified the manuscript. SR, AD, MCK have finalized the manuscript.

Conflict of Interests

The authors declare that there is no conflict of interest.

References

Anonymous, 2015. *Preliminary overview of the genus Anguilla*. UNEP-WCMC, Cambridge. Pp. 14.

Aoyama J. 2009. Life History and Evolution of Migration in Catadromous Eels (Genus *Anguilla*). Aqua-BioScience Monographs 2(1): 1–42.

Ardura A, Pola IG, Ginuino I, Gomes V and Garcia-Vazquez E. 2010. Application of barcoding to Amazonian commercial fish labelling. Food Research International 43(5): 1549-1552.

Arunachalam M and Sankaranarayanan A. 2000. Some economically important and cultivable in Gadana river, Western Ghats. In: Ponniah AG and Gopalakrishnan A (ed.), *Endemic fish diversity of Western Ghats*. NBFGR-NATP Publication, Lucknow, India. Pp: 244-246.

Bhattacharjee MJ, Laskar BA, Dhar B and Ghosh SK. 2012. Identification and Re-Evaluation of Freshwater Catfishes through DNA Barcoding. PLoS ONE 7(11): 1-7.

Bhattacharjya BK, Sugunan VV and Chaudhury M. 2000. Threatened fishes of Assam. In: Ponniah AG and Sarkar UK (ed.), *Fish diversity of North East India*. National Bureau of Fish Genetic Resources, Lucknow, India. Pp: 75-79.

Deka TK, Goswami MM and Kakati M. 2005. Causes of Fish Depletion–A Factor Analysis Approach. NAGA, WorldFish Center Newsletter 28(1): 37–42.

Dey A, Verma R, Singh M and Barat S. 2016. Evolutionary and taxonomic relationships in loach (Genus: *Botia*) through molecular characterization in a river of Terai region of West Bengal , India. European Journal of Biotechnology and Bioscience 3(9): 12-17.

Dubois A. 2003. The relationships between taxonomy and conservation biology in the century of extinctions. Comptes Rendus Biologies 326: 9-21.

Fricke R, Eschmeyer WN and van der Laan R (ed.). 2020. Eschmeyer's catalog of fishes: genera, species, references.
http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

Ghosh SK. 2012. *A text book on DNA barcoding*. Vol. 1, Books space, Kolkata, India.

Goswami UC, Basistha SK, Bora D, Shyamkumar K, Saikia B and Changsan K. 2012. Fish diversity of North East India, inclusive of the Himalayan and Indo Burma biodiversity hotspots zones: A checklist on their taxonomic status, economic importance, geographical distribution, present status and prevailing threats. International Journal of Biodiversity and Conservation 4(15): 592–613.

Groombridge B and Jenkins M. 1998. Freshwater biodiversity: a preliminary global assessment. World conservation monitoring centre 8: 104.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic. Acids. Symposium Series 41:95-98.

Hanzen C, Lucas MC, O'Brien G, Downs CT, Willows-Munro S. 2020. African freshwater eel species (*Anguilla* spp.) identification through DNA barcoding. Marine and Freshwater Research 71, 1543-1548.

Hebert PDN, Cywinska A, Ball S and de Waard J. 2003. *Biological identifications through DNA barcodes*. Proceedings. Biological sciences. Pp: 313–21.

Hoang DT, Chernomor O, Haeseler A, Von Minh BQ, and Vinh LS. 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35(2): 518–522.

Hossain MA, Islam MN, Hosain SH, Khan MFA and Khalequzzaman, SM. 2007. Status and potentials of eel fisheries in Bangladesh, Journal of Soil and Nature 1(3): 46-51.

Hubert N, Torrico JP, Bonhomme F and Renno JF. 2008. Species polyphyly and mtDNA introgression among three Serrasalmus sister-species. Molecular Phylogenetics and Evolution 46(1): 375-381.

Jacoby D, Harrison IJ and Gollock M. 2014. *Anguilla bengalensis*. The IUCN Red List of Threatened Species. https://www.iucnredlist.org/species/61668607/15501445

Jansen G, Devaere S, Weekers PHH and Adriaens D. 2006. Phylogenetic relationships and divergence time estimate of African anguilliform catfish (Siluriformes: Clariidae) inferred from ribosomal gene and spacer sequences. Molecular Phylogenetics and Evolution 38(1): 65–78.

Jayaram KC. 2010. *The fresh water fishes of Indian region*, Narendra Publishing House, New Delhi, India.

Khedkar GD, Jamdade R, Naik S, David L and Haymer D. 2014. DNA barcodes for the fishes of the Narmada, one of India's longest rivers. PloS One 9(7): 1-10.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.

Kumar S, Stecher G and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33(7): 1870–1874.

Kundu S, Chandra K, Tyagi K, Pakrashi A and Chandra K. 2019. DNA barcoding of freshwater fishes from Brahmaputra River in Eastern Himalaya biodiversity hotspot. Mitochondrial DNA Part B 4(2): 2411–2419.

Lakra WS, Singh M, Goswami M, Gopalakrishnan A, Lal KK, Mohindra, V, Sarkar UK, Punia PP, Singh KV, Bhatt JP and Ayyappan S. 2016. DNA barcoding Indian freshwater fishes. Mitochondrial DNA Part A 27(6): 4510–4517.

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishnan A, Singh KV, Ward RD and Hebert P. 2011. DNA barcoding Indian marine fishes. Molecular Ecology Resources 11(1): 60–71.

Leander NJ, Shen K-N, Chen R-T and Tzeng W-N. 2012. Species Composition and Seasonal Occurrence of Recruiting Glass Eels (*Anguilla* spp.) in the Hsiukuluan River, Eastern Taiwan. Zoological Studies 51(1): 59-71.

Menon AGK. (1992). *The fauna of India*. Vol. 4. Zoological survey of India. Kolkata, India.

Miller SA, Dykes DD and Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acid Research 16(3):1215.

Pandey PK, Singh YS, Tripathy PS, Kumar R, Abujam SK and Parhi J. 2020. DNA barcoding and phylogenetics of freshwater fish fauna of Ranganadi River, Arunachal Pradesh. Gene 754(1): 144860.

Peninal S, Subramanian J, Elavarasi A and Kalaiselvam M. 2017. Genetic Identification of Marine Eels through DNA Barcoding from Parangipettai Coastal Waters. Genomics Data 11: 81-84.

Ponniah AG and Sarkar UK. 2000. *Fish biodiversity of north east India*. National Bureau of Fish Genetic resources, Lucknow, India.

Rahman M, Norén M, Mollah AR and Kullander SO. 2019. Building a DNA barcode library for the freshwater fishes of Bangladesh. Scientific Reports 9:9382.

Saikia K and Ahmed R. 2012. Wetland fish biodiversity of Majuli river island (India) and their medicinal values. Clarion 1(2): 81–86.

Sarkar UK, Dabas A, Khan GE, Dubey VK, Kumar RR, Mishra AK, Pal A, Singh SP and Jena JK. 2013. Redescription, new distribution record, DNA sequence and length-weight relationship of the Eel-loach Pangio pangia (Cypriniformes: Cobitidae) in the River Ganges Basin, India. UNED Research Journal 5(1): 1659–4266.

Sen N. 2000. Occurance, distribution and status of diversified fish fauna of North East India. In: Ponniah AG and Sarkar UK (ed.), *Fish diversity of North East India*. National Bureau of Fish Genetic Resources, Lucknow. Pp: 31-48.

Shagufta. 2012. *Fish anatomy*. Vol: 1, APH Publishing Corporation, New Delhi, India.

Smith PJ, Veagh SMM and Steinke D. 2008. DNA barcoding for the identification of smoked fish products. Journal of Fish Biology 72: 464-471.

Stein FM, Wong JC, Sheng V, Law CS, Schroder B and Baker DM. 2016. First genetic evidence of illegal trade in endangered European eel (*Anguilla anguilla*) from Europe to Asia. Conservation Genetics Resources 8: 533–537.

Talwar PK and Jhingran AG. 1991. *Inland fishes of India and adjacent countries* Vol.1-2, Oxford & IBH Publishing Company Pvt. Ltd, New Delhi, Pp.1-1158.

Trifinopoulos J, Nguyen L, Haeseler A, Von and Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44: 232–235.

Viswanath W, Lakra WS and Sarkar UK. 2007. *Fishes of North East India*, Vol.1. National Bureau of Fish Genetic Resources, Lucknow, India.

Ward RD, Zemlak TS, Innes BH, Last P R and Hebert PDN. 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1462): 1847–1857.

Xing X, Yuan F, Huang M and Xiong X. 2020. Exploring the possible reasons for fish fraud in China based on results from monitoring sardine products sold on Chinese markets using DNA barcoding and real time PCR. Food Additives and Contaminants: Part A 37(2): 193–204.